Effects of combination of linguistic and musical pitch experience on subcortical pitch encoding

Akshay Raj Maggu[a], Patrick C.M. Wong[a,b,c], Mark Antoniou[d], Oliver Bones[e], Hanjun Liu[f,g,*], Francis C.K. Wong[h,**]

[a] Department of Linguistics and Modern Languages, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong, SAR, China
[b] Brain and Mind Institute, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong, SAR, China
[c] The Chinese University of Hong Kong-Utrecht University Joint Center for Language, Mind and Brain, Shatin, N.T., Hong Kong, SAR, China
[d] The MARCS Institute for Brain, Behaviour and Development, Western Sydney University, Locked Bag 1797, Penrith, NSW 2751, Australia
[e] Acoustics Research Centre, School of Computing, Science and Engineering, University of Salford, Salford, M5 4WT, UK
[f] Department of Rehabilitation Medicine, The First Affiliated Hospital, Sun Yat-sen University, Guangzhou, 510080, China
[g] Guangdong Provincial Key Laboratory of Brain Function and Disease, Zhongshan School of Medicine, Sun Yat-sen University, Guangzhou, 510080, China
[h] Division of Linguistics and Multilingual Studies, Nanyang Technological University, 639798, Singapore

ARTICLE INFO

Keywords:
Experience-dependent plasticity
Frequency following response
Lexical tones
Music perception

ABSTRACT

Musical experience and linguistic experience have been shown to facilitate language and music perception. However, the precise nature of music and language interaction is still a subject of ongoing research. In this study, using subcortical electrophysiological measures (frequency following response), we seek to understand the effect of interaction of linguistic pitch experience and musical pitch experience on subcortical lexical and musical pitch encoding. We compared musicians and non-musicians who were native speakers of a tone language on subcortical encoding of linguistic and musical pitch. We found that musicians and non-musicians did not differ on the brainstem encoding of lexical tones. However, musicians showed a more robust brainstem encoding of musical pitch as compared to non-musicians. These findings suggest that a combined musical and linguistic pitch experience affects auditory brainstem encoding of linguistic and musical pitch differentially. From our results, we could also speculate that native tone language speakers might use two different mechanisms, at least for the subcortical encoding of linguistic and musical pitch.

1. Introduction

Pitch is an important dimension that is relevant to both language and music perception (Plack, Oxenham, & Fay, 2005). For language, pitch is involved in signaling linguistic contrasts such as lexical tone and intonation (Ladefoged, 2003). For music, pitch is one of the central dimensions for arranging musical elements in a systematic manner (Patel, 2010). Given the important roles of pitch in both language and music, one of the intriguing questions is how the mechanisms of language and music perception interact. Previous studies show that musical experience facilitates linguistic perception (Alexander, Wong, & Bradlow, 2005; Gottfried & Riester, 2000; Gottfried, 2007; Gottfried, Staby, & Ziemer, 2004; Lee & Hung, 2008; Wong & Perrachione, 2007) and similarly,

* Corresponding author. Department of Rehabilitation Medicine, The First Affiliated Hospital, Sun Yat-sen University, Guangzhou, 510080, China.
** Corresponding author. Division of Linguistics and Multilingual Studies, Nanyang Technological University, 639798, Singapore.
E-mail addresses: lhanjun@mail.sysu.edu.cn (H. Liu), franciswong@ntu.edu.sg (F.C.K. Wong).

https://doi.org/10.1016/j.jneuroling.2018.05.003
Received 25 October 2017; Received in revised form 12 March 2018; Accepted 6 May 2018

language experience facilitates music perception (Bidelman, Gandour, & Krishnan, 2011). However, recently, it has been found that the facilitation effect of musical experience on language perception is not straightforward (Cooper & Wang, 2012), especially when multiple types of pitch experiences (language or music) are involved. In the current study, we aimed to further understand the relationship between musical experience and linguistic processing, by comparing tone-language-speakers with and without musical pitch experience, on frequency following responses (FFR) elicited from lexical and musical pitch stimuli.

1.1. Relationship between music and language: behavioral studies

Several behavioral studies have found that musical experience enhances the perception of lexical tones (Alexander et al., 2005; Gottfried & Riester, 2000; Gottfried, 2007; Gottfried et al., 2004; Lee & Hung, 2008; Wong & Perrachione, 2007). For example, Gottfried and Riester (2000) found that individuals with music major had better identification for Mandarin tones than non-majors. Further, Gottfried et al. (2004) revealed that musicians discriminated (same/different) Mandarin lexical tones more accurately than non-musicians. Alexander et al. (Alexander et al., 2005) reported that English musicians were better at both discrimination and identification of lexical tones in terms of accuracy and reaction times. In sum, these findings reflect a considerable overlap of language and music, suggestive of a common perceptual substrate for the two.

While there are studies (Alexander et al., 2005; Gottfried & Riester, 2000; Gottfried, 2007; Gottfried et al., 2004; Lee & Hung, 2008; Wong & Perrachione, 2007) conducted on non-native speakers with (and without) musical experience revealing the overlap of music and language perception, there are also studies (Lee & Lee, 2010; Lee, Lee, & Shr, 2011) conducted on native tone language speakers (with musical experience) leading to inconclusive results. Lee and colleagues found a lack of correlation between language and music perception in both tone (Lee & Lee, 2010; Lee et al., 2011) and non-tone language speakers (Lee & Hung, 2008; Lee, Lekich, & Zhang, 2014). These findings led those authors to conclude that a lack of association between lexical and musical tone identification could be due to fundamental differences in the internal category structure of lexical and musical tones and whether or not they serve a linguistic function. Identification of musical pitch solely depends on pitch but in the case of lexical tones, other cues such as amplitude and duration also play important roles (Lee et al., 2011). However, the lack of correlation between language and music perception in their studies (Lee & Hung, 2008; Lee & Lee, 2010; Lee et al., 2011) does not allow drawing definite conclusions on the language-music association (Lee et al., 2011). Their studies reveal that in the presence of more than one type of experience, the language-music interaction gets more complex. Though their studies confirm that language-music interaction is not straightforward, the nature of interaction is still unclear.

In order to understand the interaction between language and music, Cooper and Wang (2012) compared tone-language-speaking musicians and non-musicians on their abilities to learn lexical pitch in a tone-word learning paradigm. They compared the Thai- and English-speaking musicians and non-musicians on Cantonese tone identification abilities. It was predicted that Thai musicians (TM) would identify Cantonese tones most accurately due to their combined tone language and music experience (two types of experience), followed by Thai non-musicians (TNM) who would not differ from English musicians (EM; both have one type of experience), but both groups were expected to outperform English non-musicians (ENM; no tone experience). This sequence of predictions was consistent with the findings of Wong and Perrachione (2007) who had previously observed that English musicians outperformed non-musicians in the learning of novel words differentiated by lexical tone contrasts. Thus, musical experience should enhance lexical tone perception abilities. However, Cooper and Wang (2012) found that combined language and music experience did not exert an additive effect on tone identification. In fact, Thai musicians did not exhibit the expected advantage over Thai non-musicians or English musicians, but rather, both Thai non-musicians and English musicians showed more accurate tone-word identification compared to Thai musicians. As expected, English non-musicians performed worst. Cooper and Wang (2012) accounted for these findings by explaining that due to musical experience, TM may make elaborate pitch mappings (as in the case of music) when learning other pitch contours. As a result, confusion may arise between language and music for TM resulting in poorer scores than EM (who do not possess lexical tone experience) or TNM (who lack musical experience). By extension, TNM, who lack the kind of pitch acuity possessed by the TM, would have learnt the pitch contours using principles of learning a tone language, thus resulting in better performance than TM.

1.2. Relationship between music and language: brainstem electrophysiological studies

Quite recently, pitch processing has been studied using FFR, an auditory evoked potential generated predominantly at the level of inferior colliculi of the brainstem (Krishnan, Xu, Gandour, & Cariani, 2005) that is also proposedly modulated from the cortex via corticofugal pathways (Chandrasekaran & Kraus, 2010). In addition, the auditory brainstem could be influenced by linear predictive coding (Chandrasekaran, Skoe, & Kraus, 2014) in a continuous online modulation loop (Chandrasekaran, Hornickel, Skoe, Nicol, & Kraus, 2009; Kraus & Chandrasekaran, 2010) that involves co-operation of both cortex and inferior colliculi. According to continuous online modulation model, cortex predicts the incoming input from the brainstem (inferior colliculi) and if there is a match between the two, the representation is more robust throughout the central auditory system. FFR represents the phase-locking abilities of the auditory system and thus, has been used as a metric of neural plasticity following language experience (Krishnan, Gandour, Bidelman, & Swaminathan, 2009; Krishnan et al., 2005; Swaminathan, Krishnan, & Gandour, 2008) and musical training (Wong, Skoe, Russo, Dees, & Kraus, 2007). Wong et al. (2007) observed more faithful FFR encoding of natural Mandarin tones in English musicians than non-musicians. Bidelman, Gandour, and Krishnan (2009) investigated cross-domain effects of music and language experience by studying brainstem encoding of synthetically generated musical and lexical tone contours. They found that both Chinese non-musicians and English musicians encoded both musical and lexical tones more robustly compared to English non-musicians. Importantly, Chinese non-musicians and English musicians did not differ on either musical or lexical tone encoding. Further, Bidelman et al. (2011) revealed that both Chinese non-musicians and English musicians shared enhanced brainstem encoding for musical pitch. However, only musicians (but not Chinese listeners) exhibited a perceptual advantage in pitch discrimination tasks. The lack of perceptual benefits in Chinese speakers was attributed to the fact that musical pitch was behaviorally irrelevant to them as non-musicians.

Taken together, these studies on subcortical encoding of pitch reveal that lexical tones and music share a common sensory-level perceptual substrate. However, as with most behavioral studies (Alexander et al., 2005; Gottfried & Riester, 2000; Gottfried, 2007; Gottfried et al., 2004; Lee & Hung, 2008; Lee et al., 2014) examining the effect of musical experience on lexical tone perception, these brainstem electrophysiological studies (Bidelman et al., 2009, 2011; Wong et al., 2007) also focused on investigating the effect of a single type of experience (linguistic or musical) on lexical and/or musical pitch encoding. In the current study, we investigated the interactive effects of lexical and musical pitch experience on subcortical encoding of lexical and musical pitch stimuli. We compared tonal musicians and tonal non-musicians on the subcortical encoding of lexical and musical pitch, using FFR.

If both musical and linguistic pitches are processed in a similar manner in tone language speakers, we would expect a stronger corticofugal and continuous online modulation in tonal musicians due to the combined experience they possess as compared to the tonal non-musicians, leading to an enhanced brainstem encoding in the musicians than non-musicians. In other words, the effects of musical and linguistic experience would turn out to be additive. On the other hand, if the mechanisms underlying linguistic and musical pitch processing are different in tone language speakers, we would expect similar magnitude of corticofugal and continuous online modulation for tonal musicians and tonal non-musicians for lexical tones. This would result in tonal musicians having similar brainstem encoding for lexical tones but enhanced brainstem encoding for musical pitch in musicians as compared to tonal non-musicians.

We analyzed our subjects' data to compare the groups across six FFR measures, namely, stimulus-to-response correlation, pitch strength, pitch error, signal-to-noise ratio, peak F0 amplitude, and root-mean-square amplitude. These FFR measures have been found to be sensitive in evaluating timing, periodicity, and the spectral envelop information of the brainstem encoding (Liu, Maggu, Lau, & Wong, 2015; Skoe & Kraus, 2010; 2013; Song, Banai, & Kraus, 2008; Wong et al., 2007). Out of these measures, three measures (stimulus-to-response correlation, pitch strength, pitch error) tap the fidelity of the pitch of the FFR while the other three measures (signal-to-noise ratio, peak F0 amplitude, root-mean-square amplitude) tap the magnitude of the FFR signal. Our main aim was to fully understand whether or not there were any combined effects (pitch related and/or magnitude related) of linguistic and musical experience on brainstem encoding of lexical and musical tones. See section 2.3.3 for a detailed description of the measures.

2. Materials and methods

2.1. Participants

All participants were native speakers of Hong Kong Cantonese with peripheral hearing sensitivity within 25 dB HL at 0.5–4 kHz, no history of middle ear pathology, and no obvious anatomical/neurological defects. Participants with 6 years or more of formal musical training on any musical instrument were included in the musician group while those with less than 3 years of formal musical training were considered to be non-musicians. We did not recruit participants outside of this range in order to define a more distinctive musicians and non-musician group. The Joint Chinese University of Hong Kong – New Territories East Cluster Clinical Research Ethics Committee approved the study.

For collecting FFRs with lexical tones, 30 native speakers (9 males) of Hong Kong Cantonese including 15 musicians (3 males, mean age: 21.1 years) and 15 non-musicians (6 males, mean age: 22.3 years) were recruited. For collecting FFRs with musical stimuli, 30 native speakers (12 males) of Hong Kong Cantonese including 15 musicians (6 males, mean age: 21.3 years) and 15 non-musicians (6 males, mean age: 20.9 years) were recruited.

2.2. Stimuli

2.2.1. Lexical tone stimuli

The stimuli for studying lexical tone encoding in the brainstem consisted of the syllable/ji/recorded with the six lexical tones of Cantonese making six unique words:/ji1/‘doctor’,/ji2/‘chair’,/ji3/‘meaning’,/ji4/‘son’,/ji5/‘ear’ and/ji6/‘justice’. These stimuli have been used in past research (Liu et al., 2015). The stimuli were intensity normalized to 75 dB SPL and time-normalized to 175 ms using Praat (Boersma & Weenink, 2001). Fig. 1 shows the F0 contours of the six lexical tones of Cantonese (F0 ranges: 135–146 Hz, 105–134 Hz, 120–124 Hz, 85–99 Hz, 98–113 Hz, 98–106 Hz).

2.2.2. Musical pitch stimuli

Stimuli for studying musical pitch encoding were cello stimuli adapted from Musacchia, Sams, Skoe, and Kraus (2007) that were manipulated for fundamental frequency (F0) to match the registers of high (Tone 1) and low-level tones (Tone 6) of Cantonese (Liu et al., 2015).

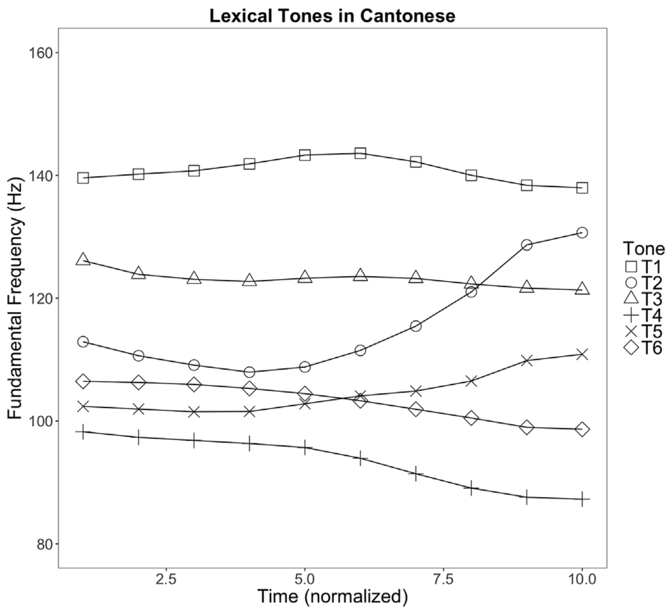

Fig. 1. F0 contours of the six Cantonese lexical tones (F0 ranges: T1: 135–146 Hz, T2: 105–134 Hz, T3: 120–124 Hz, T4: 85–99 Hz, T5: 98–113 Hz, T6: 98–106 Hz).

2.3. *Procedure*

2.3.1. *Stimuli presentation*

Participants heard a total of 3000 sweeps for each stimulus in alternating polarity in order to minimize stimulus artifacts. Stimuli were presented in their right ear via insert earphones (Compumedics 10Ω) at 81 dB SPL using the Audio CPT module of STIM2 (Compumedics, USA). Inter-stimulus (offset to onset) interval was jittered from 74 to 104 ms (Liu et al., 2015; Maggu, Liu, Antoniou, & Wong, 2016; Wong et al., 2007) and the order of stimulus presentation was randomized across participants. Participants were asked to relax and ignore the stimuli.

2.3.2. *Data acquisition and pre-processing*

Continuous electrophysiological data were collected using Ag/AgCl electrodes at Cz (active) referenced to linked M1 and M2 (linked mastoids) with lower forehead as ground and the inter-electrode impedances maintained at ≤ 1 kΩ. The data were collected at a sampling rate of 20,000 Hz using a Synamps RT amplifier (Compumedics, El Paso, TX). Offline data pre-processing that consisted of artifact rejection (± 35 μV), filtering (80–5000 Hz; 6 dB roll off), epoching, and averaging was carried out using Curry 7.05 (Compumedics, El Paso, TX). Three FFR recordings with more than 10% of rejected sweeps (i.e., > 300 rejections) were removed and not included in further analyses.

2.3.3. *FFR data analysis procedures*

As we were interested in studying the phase-locking of the auditory brainstem, we further band-pass filtered the FFRs in the range of 80–2500 Hz to remove any contribution from slower cortical ERPs and to attenuate the high frequency component of EEG noise (Bidelman, Hutka, & Moreno, 2013; Skoe & Kraus, 2010). The data were further converted from temporal to spectral domain and processed using a sliding window Fast Fourier Transform (FFT) analysis (Liu et al., 2015; Song et al., 2008; Wong et al., 2007) where 125 Hanning-windowed overlapping bins (zero padded) were obtained across the 175-msec of brainstem response by shifting 50-msec sliding window in 1-msec steps. Pitch (F0) contours of FFR were obtained by connecting the spectral peaks (nearest to the expected stimulus frequency) from these 125 overlapping bins. The following measures were obtained to compare the musicians and non-musicians on brainstem encoding of pitch (Liu et al., 2015; Skoe & Kraus, 2010; Song et al., 2008; Wong et al., 2007).

a. Stimulus-to-Response Correlation (values ranging from -1 to + 1) is a Pearson correlation (*r*) between the pitch contour of a stimulus and its response. Stimulus-to-Response correlation reflects the ability of brainstem in recapitulation of the original input signal. Higher values of Stimulus-to-Response correlation reflect better encoding of pitch in the brainstem.

b. Pitch Strength (values between + 1 and -1) is obtained by autocorrelation technique and is a measure of periodicity of the response. It was calculated by measuring the autocorrelation peaks from 125 bins in each FFR (Liu et al., 2015).

c. Signal-to-Noise Ratio (SNR) refers to the ratio of RMS amplitude of the response to the RMS amplitude of the pre-stimulus period (−50 ms).

d. Peak F0 amplitude (in dB, peak amplitudes in the power spectrum) was obtained by measuring the peak amplitude among the spectral peaks in the frequency range of fundamental frequency (F0).

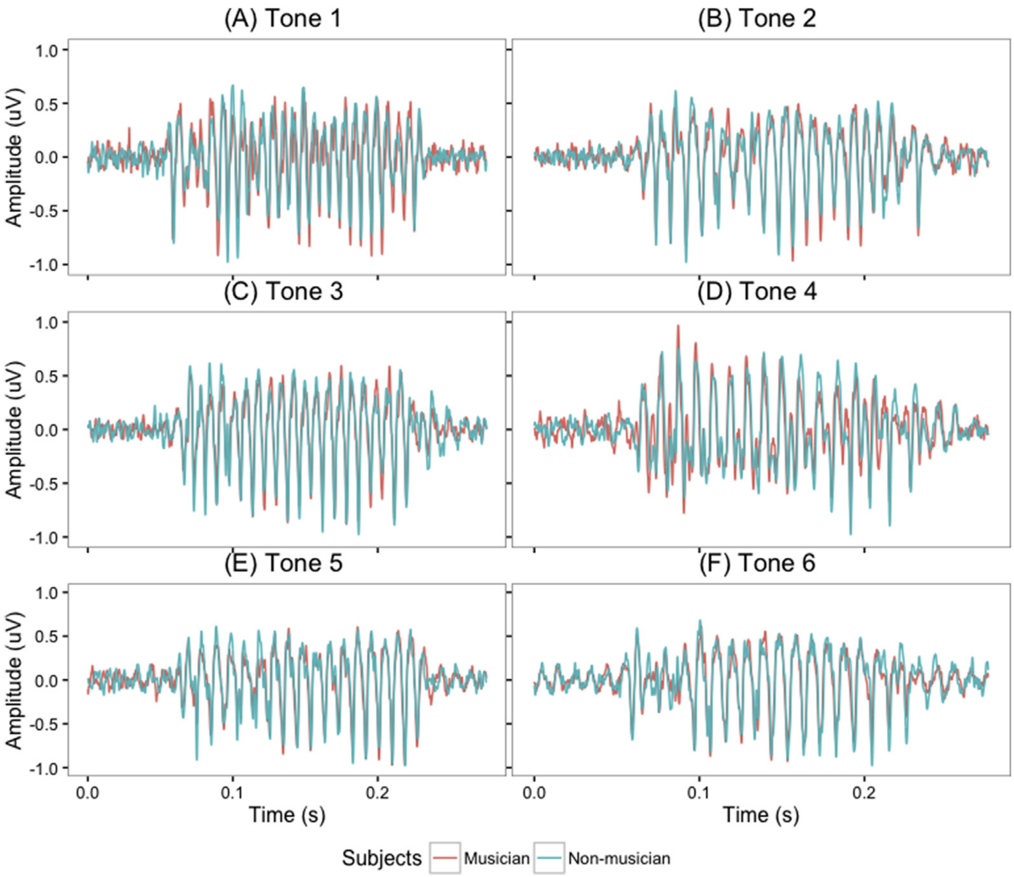

Fig. 2. Comparison of grand-averaged FFR waveforms of musicians and non-musicians across the six lexical tones of Cantonese (A–F). X-axis: Time (s), Y-axis: Amplitude (μV). Initial and final 50 ms represent pre- and post-stimulus baselines. (colour to be used). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

e. Pitch error (in Hz) is the average Euclidian distance between the stimulus pitch contour (F0) and response pitch contour (F0). The lower the pitch error, the better the pitch encoding at the brainstem.

f. Root-mean-square amplitude (RMS; in μV) refers to the magnitude of activation of the neural response across the entire FFR duration (175 ms).

3. Results

3.1. Brainstem encoding of lexical tones

Grand-averaged FFR waveforms of the musician and non-musician groups in response to each of the six Cantonese lexical tones are shown in Fig. 2. Fig. 3 shows the pitch tracking of grand-averaged FFRs of the musicians (yellow patches) and non-musicians (green patches) and their comparison to the pitch contour for each corresponding stimulus (black lines). A series of 2 (Group: Musicians vs non-musicians) × 6 (Tone: 6 lexical tones) ANOVAs were carried out for each of the brainstem measures.

a. Stimulus-to-Response Correlation. For stimulus-to-response correlation, there was no main effect of group, $F(1, 28) = 1.27$, $p = .27$, $\eta_p^2 = 0.043$, nor was there a significant interaction, $F(5, 140) = 1.31$, $p = .26$, $\eta_p^2 = 0.045$. There was a significant main effect of tone, $F(5, 140) = 17.9$, $p < .001$, $\eta_p^2 = 0.390$ (see Fig. 4A). Post-hoc analyses revealed that the stimulus-to-response correlation for Tone 3 was lower than for all other tones, and the stimulus-to-response correlation for Tone 1 was lower than for Tones 2, 4, and 5.

b. Pitch Strength. For pitch strength, there was no main effect of group, $F(1, 28) = 0.412$, $p = .53$, $\eta_p^2 = 0.015$, nor was there a significant interaction, $F(5, 140) = 0.44$, $p = .82$, $\eta_p^2 = 0.016$ (see Fig. 4B). There was a significant main effect of tone, $F(5, 140) = 8.1$, $p < .001$, $\eta_p^2 = 0.226$. Post-hoc analyses revealed that Tone 1 had lower pitch strength than Tones 2, 3, 4, 5, and 6.

c. SNR. For SNR, there was no main effect of group, $F(1, 28) = 1.59$, $p = .22$, $\eta_p^2 = 0.054$, nor was there a significant interaction, $F(5, 140) = 1.22$, $p = .30$, $\eta_p^2 = 0.042$ (see Fig. 4C). There was a significant main effect of tone, $F(5, 140) = 5.7$, $p < .001$, $\eta_p^2 = 0.168$. Post-hoc analyses revealed that Tone 1 had lower SNR than Tones 2, 3, 4, and 6, and Tone 3 had higher SNR than Tones 1,

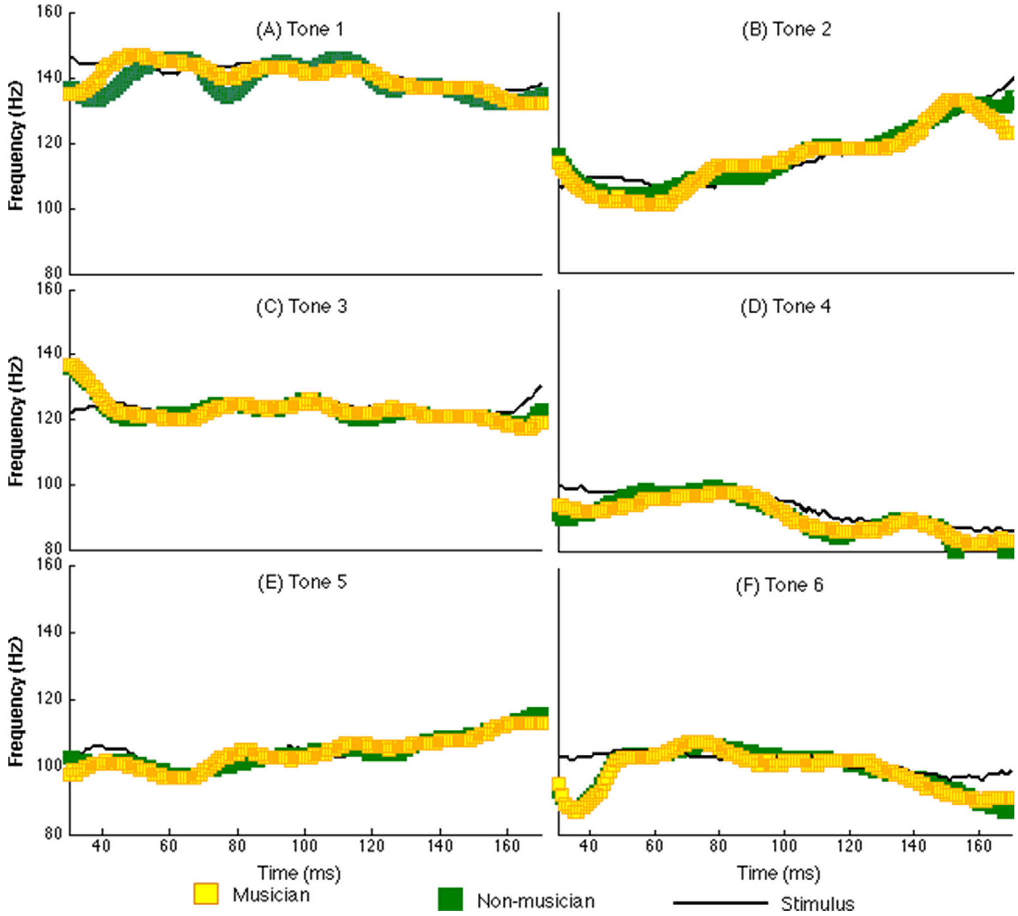

Fig. 3. Comparison of pitch (F0) tracking of grand-averaged FFRs of musicians and non-musicians across the six lexical tones (A–F) (colour to be used). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

5, and 6.

d. Peak F0 amplitude. For peak F0 amplitude, there were no significant main effects of group, $F(1, 28) = 0.59$, $p = .45$, $\eta_p^2 = 0.021$, or tone, $F(5, 140) = 0.77$, $p = .57$, $\eta_p^2 = 0.027$, nor was there a significant interaction, $F(5, 140) = 0.54$, $p = .75$, $\eta_p^2 = 0.019$ (see Fig. 4D).

e. Pitch Error. For pitch error, there were no significant main effects of group, $F(1, 28) = 0.02$, $p = .90$, $\eta_p^2 = 0.001$, or tone, $F(5, 140) = 2.1$, $p = .069$, $\eta_p^2 = 0.070$, nor was there a significant interaction, $F(5, 140) = 0.29$, $p = .91$, $\eta_p^2 = 0.010$ (see Fig. 4E).

f. RMS amplitude. For RMS amplitude, there were no significant main effects of group, $F(1, 28) = 0.98$, $p = .33$, $\eta_p^2 = 0.034$, or tone, $F(5, 140) = 1.3$, $p = .27$, $\eta_p^2 = 0.044$, nor was there a significant interaction, $F(5, 140) = 0.48$, $p = .79$, $\eta_p^2 = 0.017$ (see Fig. 4F).

In sum, the FFR results for the Cantonese lexical tones show no significant main effects or interactions involving the group factor for any of the brainstem measures, suggesting that musicians and non-musicians do not differ on brainstem encoding of lexical tones (Figs. 2–4).

3.2. Brainstem encoding of musical pitch

Fig. 5 shows the comparison between musicians and non-musicians across F0 amplitude and grand-averaged FFR waveforms for the musical pitch stimuli.

A series of 2 (Group) × 2 (Tone) ANOVAs were carried out, one for each of the brainstem measures.

a. Stimulus-to-Response Correlation. For stimulus-to-response correlation, there was a significant main effect of group, $F(1, 28) = 10.23$, $p = .003$, $\eta_p^2 = 0.268$, confirming that musicians showed more faithful encoding of pitch than non-musicians (Cello Tone 1: $M = 0.524$ vs. 0.346; Cello Tone 6: $M = 0.469$ vs 0.282, respectively), as depicted in Fig. 6A. There was no significant main effect of tone, $F(1, 28) = 3.38$, $p = .077$, $\eta_p^2 = 0.108$, or interaction, $F(1, 28) = 0.02$, $p = .90$, $\eta_p^2 = 0.001$.

b. Pitch strength. For pitch strength, there was no main effect of group, $F(1, 28) = 0.19$, $p = .67$, $\eta_p^2 = 0.007$, but there was a

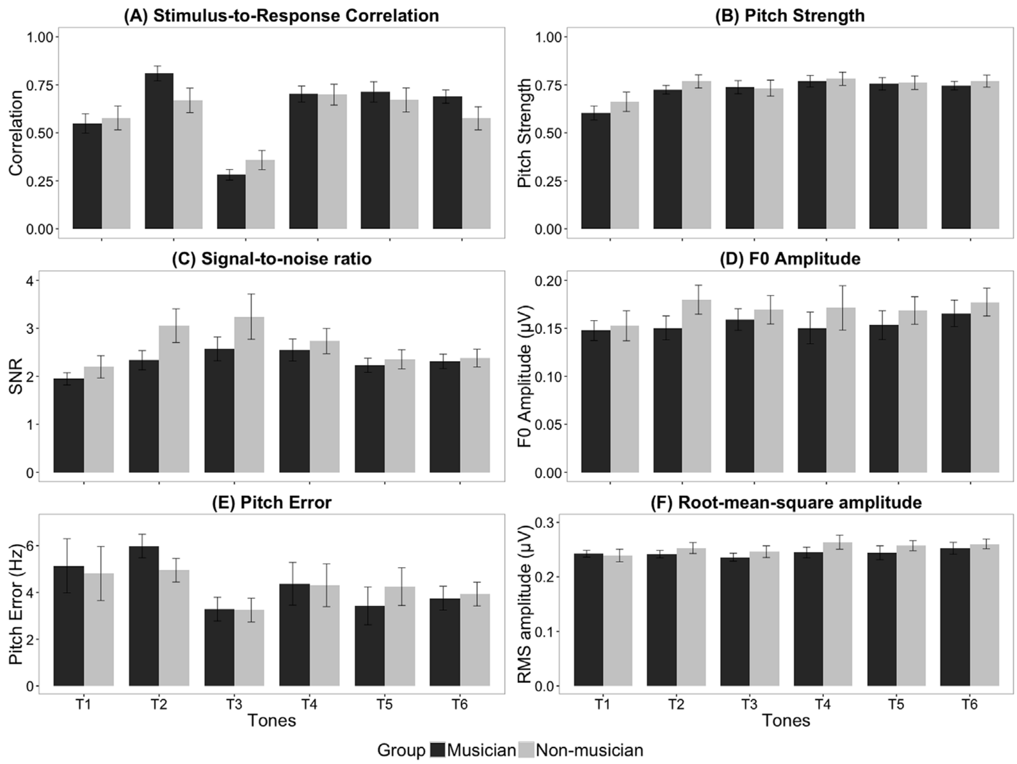

Fig. 4. Comparison of Musician and Non-Musician groups across the six lexical tones on the FFR measures: (A) Stimulus-to-response correlation (B) Pitch Strength (C) Signal-to-Noise ratio (D) F0 amplitude (E) Pitch Error (F) RMS amplitude. No significant main effect of Group (musicians vs non-musicians) was found in any of the measures (Error bars = ± SEM).

significant interaction, $F(1, 28) = 4.12$, $p = .050$, $\eta_p^2 = 0.130$. Although Fig. 6B seems to suggest that musicians showed more robust encoding of cello tone 1, but not for cello tone 6, post-hoc analyses confirmed that neither of the differences were statistically significant ($p = .32$ and $p = .75$, respectively). There was a significant main effect of tone, $F(1, 28) = 10.05$, $p = .004$, $\eta_p^2 = 0.264$, showing that overall, Tone 6 had higher pitch strength than Tone 1.

c. SNR. For SNR, there was a significant main effect of group, $F(1, 28) = 17.89$, $p < .001$, $\eta_p^2 = 0.390$, revealing that SNR was greater for musicians than non-musicians (Cello Tone 1: $M = 3.05$ vs. 1.98; Cello Tone 6: $M = 3.81$ vs. 2.76, respectively), as is shown in Fig. 6C. There was also a significant main effect of tone, $F(1, 28) = 14.94$, $p = .001$, $\eta_p^2 = 0.348$, showing that Tone 6 had a higher SNR than Tone 1. There was no significant interaction, $F(1, 28) = 0.01$, $p = .91$, $\eta_p^2 = 0.000$.

d. Peak F0 amplitude. For peak F0 amplitude, there was a significant main effect of group, $F(1, 28) = 4.81$, $p = .037$, $\eta_p^2 = 0.147$, confirming that musicians exhibited greater peak F0 amplitude than non-musicians (Cello Tone 1: $M = 0.081$ vs. 0.062; Cello Tone 6: $M = 0.087$ vs. 0.081), respectively (Fig. 5A, B and 6D). There was also a significant main effect of tone, $F(1, 28) = 7.68$, $p = .010$, $\eta_p^2 = 0.215$, showing that Tone 6 had higher peak F0 amplitude than Tone 1. There was no significant interaction, $F(1, 28) = 2.27$, $p = .14$, $\eta_p^2 = 0.075$.

e. Pitch error. For pitch error, there was no main effect of group, $F(1, 28) = 0.72$, $p = .40$, $\eta_p^2 = 0.025$, nor was there a significant interaction, $F(1, 28) = 0.41$, $p = .53$, $\eta_p^2 = 0.014$. There was a significant main effect of tone, $F(1, 28) = 8.17$, $p = .008$, $\eta_p^2 = 0.226$, showing that Tone 1 had a larger pitch error than Tone 6 (Fig. 6E).

f. RMS amplitude. For RMS amplitude, there was a significant main effect of group, $F(1, 28) = 4.93$, $p = .035$, $\eta_p^2 = 0.150$, and also a significant interaction, $F(1, 28) = 4.5$, $p = .042$, $\eta_p^2 = 0.139$ (Fig. 6F). Post-hoc t-tests revealed that musicians had higher RMS amplitude than non-musicians on cello tone 6 ($M = 0.275$ vs. $M = 0.242$, respectively), $t(28) = 2.66$, p, $= .013$, but not cello tone 1, $p = .845$ (Fig. 5C and D).

These results indicate that musicians outperformed the non-musicians on the brainstem encoding of musical pitch. Also, overall, the Tone 6 cello was encoded more robustly than Tone 1 cello.

4. Discussion

In the current study, we investigated the interactive effects of musical and tone language experience on lexical and musical pitch perception at the subcortical levels. The main finding of our study is that musical experience does not significantly enhance the lexical tone encoding of Cantonese musicians compared to Cantonese non-musicians. However, we found that the Cantonese musicians

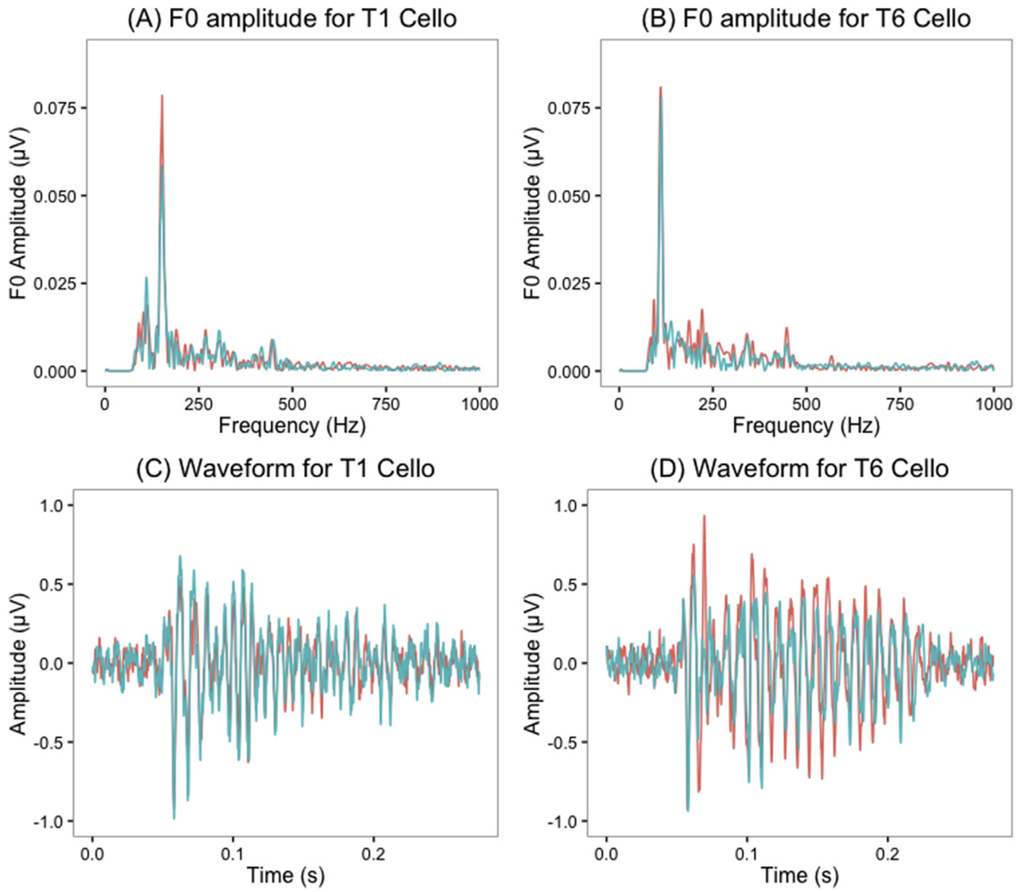

Fig. 5. Comparison of Musician and Non-Musician groups on F0 amplitude (A and B) and RMS amplitude (C and D). Musicians have significantly better F0 amplitude (p = .037) on Cello T1 (A) and significantly better RMS amplitude (p = .04) on Cello T6 (D) than Non-Musicians. (colour to be used). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

performed significantly better than Cantonese non-musicians on musical pitch encoding.

Previous findings suggest that musical experience enhances brainstem encoding of lexical tones (Bidelman et al., 2011; Wong et al., 2007), and lexical tone experience enhances brainstem encoding of musical pitch (Bidelman et al., 2011). By exploring the combined effects of language and music experience, the present study has revealed that musical experience does not further enhance lexical tone encoding in individuals with tone language experience. These findings can be supported by the behavioral findings of Cooper and Wang (2012) who found a no significant difference between Thai musicians and Thai non-musicians on lexical tone perception. One might speculate that these findings stem from different mechanisms involved in learning a tone language versus learning music. The current findings are also consistent with the behavioral findings of (Mok & Zuo, 2012) who found no facilitation effect of musical experience on AX discrimination of Cantonese tone-merging categories when they compared Cantonese musicians and non-musicians. They suggest that native tone language speakers may be using different mechanisms for perceiving lexical and musical pitch. Taken together, the current findings reveal that the effects of combination of different types of experience (language and music) are not simply additive.

4.1. Neurophysiological explanation of the current findings

From a neurophysiological standpoint, there are a few possibilities that can account for the current findings. First, brainstem encoding is proposedly modulated from the cortex via corticofugal pathways (Bajo, Nodal, Moore, & King, 2010; Hairston, Letowski, & McDowell, 2013; Krizman, Marian, Shook, Skoe, & Kraus, 2012; Parbery-Clark, Skoe, & Kraus, 2009). Both Cantonese musicians and non-musicians with similar degrees of tone language experience, listening to lexical tones might have been equally influenced by corticofugal modulation. In other words, the absence of a significant difference between the two groups on lexical tone brainstem encoding could be speculated to be a result of similar degrees of corticofugal modulation. On the other hand, Cantonese musicians that differed from the non-musicians only on the degree of musical experience could have an increased corticofugal modulation for the musical pitch, leading to enhanced brainstem encoding. Another more plausible explanation for the current neural findings could be from the linear predictive tuning model (Chandrasekaran et al., 2014) that proposes both corticofugal and local modulation at the

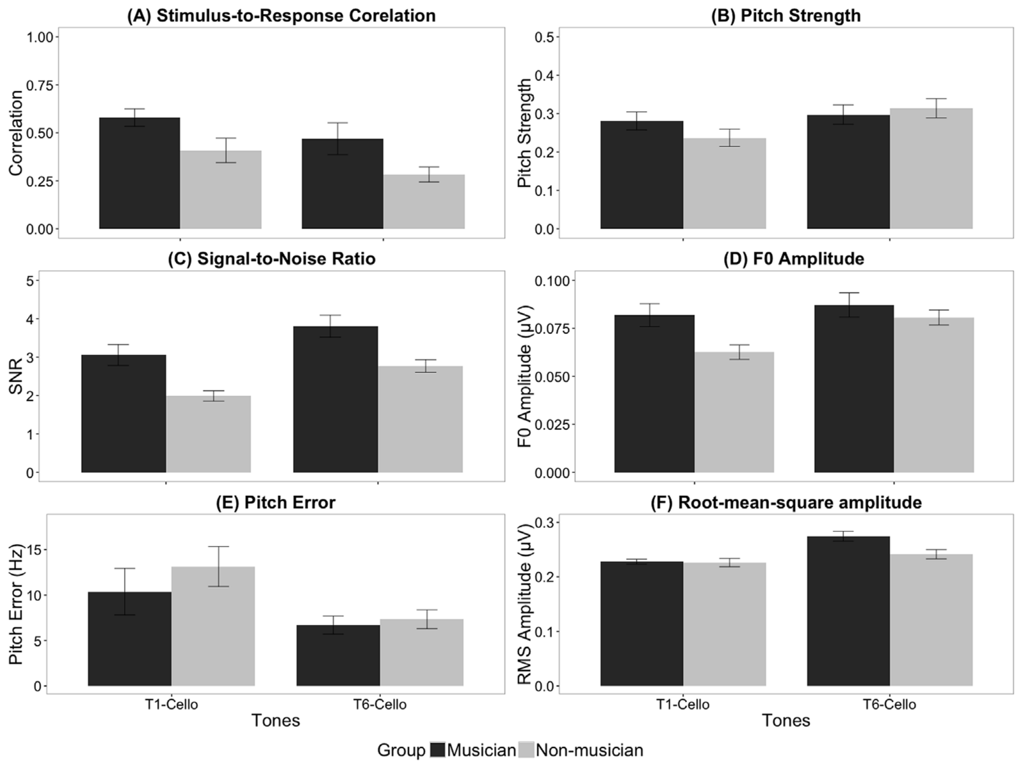

Fig. 6. Comparison of Musician and Non-Musician groups across the two musical stimuli (Cello T1 and Cello T6) on FFR measures: (A) Stimulus-to-response correlation, (B) Pitch Strength, (C) Signal-to-Noise ratio, (D) F0 amplitude, (E) Pitch Error, and (F) RMS amplitude. Panel (A), (C), (D) and (F): Significant main effect of Group (musicians outperformed the non-musicians); Panel (B) and (E): No significant main effect of Group. (Error bars = ± SEM).

level of the inferior colliculi (Krishnan, Gandour, & Bidelman, 2010; Krishnan, Gandour, et al., 2009; Krishnan et al., 2005) play crucial roles in brainstem encoding. According to this model, there is a continuous online modulation of brainstem encoding by the cortex via corticofugal pathways (Chandrasekaran et al., 2009; Kraus & Chandrasekaran, 2010) with processes in the inferior colliculi still active. The cortical level predicts the input from subcortical levels and when there is a match between the two, the representation is more robust throughout the central auditory system. Listening to lexical tones, both Cantonese musicians and non-musicians might have similar predictions of the incoming stimuli at the cortical level, thus leading to equally robust brainstem encoding of lexical tones in Cantonese musicians and non-musicians. On the other hand, listening to musical stimuli, musicians might predict the incoming musical pitch more accurately than non-musicians as a result of which they might show a more robust brainstem encoding than non-musicians.

The current findings can also be explained via functional relevance of the lexical/musical pitch stimuli used in the study. It is possible that musical pitch stimuli might have been more functionally relevant to Cantonese musicians than Cantonese non-musicians that would have led the Cantonese musicians towards an enhanced musical pitch encoding as compared to Cantonese non-musicians. This explanation is in line with the findings of Bidelman et al. (2011) who reported (from their behavioral experiment) musicians to have more perceptual advantages to process musical pitch as compared to tone language speakers mainly because the musical stimuli were functionally more irrelevant to them than the tone language speakers.

4.2. Neural encoding of different pitch contours

In the lexical tones FFR, we found that on almost all the measures of FFR, Tone 2 (high rising) showed a slightly better encoding than the rest of the lexical tones. These findings are consistent with the literature (Krishnan, Xu, Gandour, & Cariani, 2004) that report that a rising pitch contour is better encoded in the brainstem than other pitch contours. These findings can also be supported by the electrophysiological findings that report selectivity to rising tonal stimuli for cochlear microphonics (Shore & Cullen, 1984), eight nerve compound action potentials (Shore & Nuttall, 1985), and responses of the ventral cochlear units (Shore, Clopton, & Au, 1987). These findings report that there is more displacement of the cochlear partition for a rising tone (Shore & Cullen, 1984) that possibly leads to more synchronous activity at the level of the eight nerve (Shore & Nuttall, 1985) potentially leading to an enhanced magnitude of the compound action potential. In the musical pitch FFR, on almost all the measures, we found that Tone 6-cello (low level) showed better brainstem encoding as compared to Tone 1-cello (high level). Tone 6 is slightly better than Tone 1 on the lexical tones FFR. Till date, as far as we know, there are no reports that explain why a low level tone could be better than a high level tone either on brainstem encoding or in perception. Though more investigation is needed, we believe that this could also be influenced by the acoustic characteristics of the speech stimuli and listeners' abilities.

One of the caveats in the current study could be a possible “saturation effect” due to both the subject groups being the speakers of a tone language. However, since there is a lack of ceiling effect in our FFR measures, we argue that our data cannot be explained by a saturation scenario. In other words, if there are additive effects of language and musical experience, they should be visible in the FFR magnitude or pitch measures.

4.3. Limitations and future directions

Though the current electrophysiological study addresses a comparison between tonal musicians and tonal non-musicians on brainstem encoding of lexical and musical tones, one of the limitations of the current study could be the absence of non-tone musician and non-musician groups. The presence of the non-tone language groups could have led to a fuller understanding of interactive effects of language and music. In addition, recent reports (Yu & Zhang, 2018) suggest that there is a lack of correlation between FFR measures and lexical tone perception. In the light of these reports, it becomes necessary to conduct behavioral testing of lexical and musical tone perception along with FFR evaluation to understand the interaction of language and music in both behavior and in neurophysiology. Further, given the possibilities of a lexical-semantic confound and dependence on the acoustic characteristics of the consonant and vowel in the case of use of speech stimuli, future studies in this direction could consider using non-speech stimuli such as iterated rippled noise (Krishnan, Swaminathan, & Gandour, 2009; Yu & Zhang, 2018).

5. Conclusion

In the present study, we investigated whether the effects of linguistic and musical pitch experience are additive at the level of the brainstem, using FFR. We found that there is no additional advantage towards subcortical encoding of lexical tones when more than one type of experience (musical and linguistic) is present. However, we found that Cantonese musicians performed slightly better than the Cantonese non-musicians on musical pitch encoding, probably due to different mechanisms being used by the native tone language speakers in perceiving lexical and musical pitch.

Conflicts of interest

There are no competing interests.

Acknowledgements

This research was supported by the Dr Stanley Ho Medical Development Foundation, National Institutes of Health Grant Nos. R01DC008333 and R01DC013315, National Science Foundation (USA)National Science Foundation Grant No. BCS-1125144, and Research Grants Council of Hong Kong Grant Nos. 477513 and 14117514 to P.C.M.W. We thank Dr. Erika Skoe, University of Connecticut, for providing us the cello stimulus. We thank Doris Lau and Hilda Chan for their assistance in this study.

References

Alexander, J. A., Wong, P. C. M., & Bradlow, A. R. (2005). Lexical tone perception in musicians and non-musicians. *Interspeech* (pp. 397–400). . Retrieved from http://www.academia.edu/download/40624850/Lexical_tone_perception_in_musicians_and20151203-29900-c6mvit.pdf.

Bajo, V. M., Nodal, F. R., Moore, D. R., & King, A. J. (2010). The descending corticocollicular pathway mediates learning-induced auditory plasticity. *Nature Neuroscience, 13*(2), 253–260.

Bidelman, G. M., Gandour, J. T., & Krishnan, A. (2009). Cross-domain effects of music and language experience on the representation of pitch in the human auditory brainstem. *Journal of Cognitive Neuroscience, 23*(2), 425–434. https://doi.org/10.1162/jocn.2009.21362.

Bidelman, G. M., Gandour, J. T., & Krishnan, A. (2011). Musicians and tone-language speakers share enhanced brainstem encoding but not perceptual benefits for musical pitch. *Brain and Cognition, 77*(1), 1–10. https://doi.org/10.1016/j.bandc.2011.07.006.

Bidelman, G. M., Hutka, S., & Moreno, S. (2013). Tone language speakers and musicians share enhanced perceptual and cognitive abilities for musical Pitch: Evidence for bidirectionality between the domains of language and music. *PLoS One, 8*(4), e60676. https://doi.org/10.1371/journal.pone.0060676.

Boersma, P., & Weenink, D. (2001). Praat, a system for doing phonetics by computer. *Glot International, 5*(9/10), 341–345.

Chandrasekaran, B., Hornickel, J., Skoe, E., Nicol, T., & Kraus, N. (2009). Context-dependent encoding in the human auditory brainstem relates to hearing speech in noise: Implications for developmental dyslexia. *Neuron, 64*(3), 311–319. https://doi.org/10.1016/j.neuron.2009.10.006.

Chandrasekaran, B., & Kraus, N. (2010). The scalp-recorded brainstem response to speech: Neural origins and plasticity. *Psychophysiology, 47*(2), 236–246. https://doi.org/10.1111/j.1469-8986.2009.00928.x.

Chandrasekaran, B., Skoe, E., & Kraus, N. (2014). An integrative model of subcortical auditory plasticity. *Brain Topography, 27*(4), 539–552. https://doi.org/10.1007/s10548-013-0323-9.

Cooper, A., & Wang, Y. (2012). The influence of linguistic and musical experience on Cantonese word learning. *Journal of the Acoustical Society of America, 131*(6), 4756–4769.

Gottfried, T. L. (2007). Music and language learning. *Language Experience in Second Language Speech Learning,* 221–237.

Gottfried, T. L., & Riester, D. (2000). Relation of pitch glide perception and Mandarin tone identification. *Journal of the Acoustical Society of America, 108*(5), 2604.

Gottfried, T. L., Staby, A. M., & Ziemer, C. J. (2004). Musical experience and Mandarin tone discrimination and imitation. *Journal of the Acoustical Society of America, 115*(5) 2545–2545.

Hairston, W. D., Letowski, T. R., & McDowell, K. (2013). Task-related suppression of the brainstem frequency following response. *PLoS One, 8*(2), e55215.

Kraus, N., & Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nature Reviews Neuroscience, 11*(8), 599–605. https://doi.org/10.1038/nrn2882.

Krishnan, A., Gandour, J. T., & Bidelman, G. M. (2010). The effects of tone language experience on pitch processing in the brainstem. *Journal of Neurolinguistics, 23*(1),

81–95.
Krishnan, A., Gandour, J. T., Bidelman, G. M., & Swaminathan, J. (2009a). Experience dependent neural representation of dynamic pitch in the brainstem. *NeuroReport, 20*(4), 408–413. https://doi.org/10.1097/WNR.0b013e3283263000.
Krishnan, A., Swaminathan, J., & Gandour, J. T. (2009b). Experience-dependent enhancement of linguistic pitch representation in the brainstem is not specific to a speech context. *Journal of Cognitive Neuroscience, 21*(6), 1092–1105. https://doi.org/10.1162/jocn.2009.21077.
Krishnan, A., Xu, Y., Gandour, J. T., & Cariani, P. A. (2004). Human frequency-following response: Representation of pitch contours in Chinese tones. *Hearing Research, 189*(1–2), 1–12. https://doi.org/10.1016/S0378-5955(03)00402-7.
Krishnan, A., Xu, Y., Gandour, J., & Cariani, P. (2005). Encoding of pitch in the human brainstem is sensitive to language experience. *Cognitive Brain Research, 25*(1), 161–168. https://doi.org/10.1016/j.cogbrainres.2005.05.004.
Krizman, J., Marian, V., Shook, A., Skoe, E., & Kraus, N. (2012). Subcortical encoding of sound is enhanced in bilinguals and relates to executive function advantages. *Proceedings of the National Academy of Sciences, 109*(20), 7877–7881.
Ladefoged, P. (2003). The measurement of linguistic differences in pitch. In E. Herrera, & P. Martín (Eds.). *La Tonía: Dimensiones fonéticas y fonológicas* (pp. 375–402). .
Lee, C.-Y., & Hung, T.-H. (2008). Identification of Mandarin tones by English-speaking musicians and nonmusicians. *Journal of the Acoustical Society of America, 124*(5), 3235–3248. https://doi.org/10.1121/1.2990713.
Lee, C.-Y., & Lee, Y.-F. (2010). Perception of musical pitch and lexical tones by Mandarin-speaking musicians. *Journal of the Acoustical Society of America, 127*(1), 481–490.
Lee, C.-Y., Lee, Y.-F., & Shr, C.-L. (2011). Perception of musical and lexical tones by Taiwanese-speaking musicians. *Journal of the Acoustical Society of America, 130*(1), 526–535.
Lee, C.-Y., Lekich, A., & Zhang, Y. (2014). Perception of pitch height in lexical and musical tones by English-speaking musicians and nonmusiciansa. *Journal of the Acoustical Society of America, 135*(3), 1607–1615.
Liu, F., Maggu, A. R., Lau, J. C. Y., & Wong, P. C. M. (2015). Brainstem encoding of speech and musical stimuli in congenital amusia: Evidence from Cantonese speakers. *Frontiers in Human Neuroscience, 8*(1029)https://doi.org/10.3389/fnhum.2014.01029.
Maggu, A. R., Liu, F., Antoniou, M., & Wong, P. C. M. (2016). Neural correlates of indicators of sound change in Cantonese: Evidence from cortical and subcortical processes. *Frontiers in Human Neuroscience, 10*https://doi.org/10.3389/fnhum.2016.00652.
Mok, P. K. P., & Zuo, D. (2012). The separation between music and speech: Evidence from the perception of Cantonese tones. *Journal of the Acoustical Society of America, 132*(4), 2711–2720. https://doi.org/10.1121/1.4747010.
Musacchia, G., Sams, M., Skoe, E., & Kraus, N. (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proceedings of the National Academy of Sciences, 104*(40), 15894–15898. https://doi.org/10.1073/pnas.0701498104.
Parbery-Clark, A., Skoe, E., & Kraus, N. (2009). Musical experience limits the degradative effects of background noise on the neural processing of sound. *Journal of Neuroscience, 29*(45), 14100–14107.
Patel, A. D. (2010). *Music, language, and the brain*(1 edition). Oxford; New York: Oxford University Press.
Plack, C. J., Oxenham, A. J., & Fay, R. R. (2005). *Pitch: Neural coding and perception*(2005 edition). New York: Springer.
Shore, S. E., Clopton, B. M., & Au, Y. N. (1987). Unit responses in ventral cochlear nucleus reflect cochlear coding of rapid frequency sweeps. *Journal of the Acoustical Society of America, 82*(2), 471–478. https://doi.org/10.1121/1.395448.
Shore, S. E., & Cullen, J. K. (1984). Cochlear microphonic responses of the peripheral auditory system to frequency-varying signals. *American Journal of Otolaryngology, 5*(1), 34–42.
Shore, S. E., & Nuttall, A. L. (1985). High-synchrony cochlear compound action potentials evoked by rising frequency-swept tone bursts. *Journal of the Acoustical Society of America, 78*(4), 1286–1295.
Skoe, E., & Kraus, N. (2010a). Auditory brain stem response to complex sounds: A tutorial. *Ear and Hearing, 31*(3), 302–324. https://doi.org/10.1097/aud.0b013e3181cdb272.
Skoe, E., & Kraus, N. (2010b). Hearing it again and again: On-line subcortical plasticity in humans. *PLoS One, 5*(10), e13645. https://doi.org/10.1371/journal.pone.0013645.
Skoe, E., & Kraus, N. (2013). Musical training heightens auditory brainstem function during sensitive periods in development. *Frontiers in Psychology, 4*. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3777166/.
Song, J. H., Banai, K., & Kraus, N. (2008). Brainstem timing deficits in children with learning impairment may result from corticofugal origins. *Audiology and Neurotology, 13*(5), 335–344. https://doi.org/10.1159/000132689.
Swaminathan, J., Krishnan, A., & Gandour, J. T. (2008). Pitch encoding in speech and nonspeech contexts in the human auditory brainstem. *NeuroReport, 19*(11), 1163–1167. https://doi.org/10.1097/WNR.0b013e3283088d31.
Wong, P. C. M., & Perrachione, T. K. (2007). Learning pitch patterns in lexical identification by native English-speaking adults. *Applied PsychoLinguistics, 28*(04), 565–585.
Wong, P. C. M., Skoe, E., Russo, N. M., Dees, T., & Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuroscience, 10*(4), 420–422. https://doi.org/10.1038/nn1872.
Yu, L., & Zhang, Y. (2018). Testing native language neural commitment at the brainstem level: A cross-linguistic investigation of the association between frequency-following response and speech perception. *Neuropsychologia, 109*, 140–148. https://doi.org/10.1016/j.neuropsychologia.2017.12.022.